

# zmPDSwR Chapter 2

coop711

2015년 9월 10일

## Data Read In

- Listing 2.1

```
setwd("~/Dropbox/Works/Class/Data_Science/R.WD/zmPDSwR/")
uciCar <- read.table("../zmPDSwR/UCICar/car.data.csv", sep=",", header=TRUE)
```

- Listing 2.2

```
class(uciCar)
```

```
## [1] "data.frame"
```

```
summary(uciCar)
```

```
##      buying      maint      doors
## Length:1728    Length:1728    Length:1728
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##      persons      lug_boot      safety
## Length:1728    Length:1728    Length:1728
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##      rating
## Length:1728
## Class :character
## Mode  :character
```

```
dim(uciCar)
```

```
## [1] 1728    7
```

```
str(uciCar)
```

```
## 'data.frame':    1728 obs. of  7 variables:
## $ buying   : chr  "vhigh" "vhigh" "vhigh" "vhigh" ...
## $ maint    : chr  "vhigh" "vhigh" "vhigh" "vhigh" ...
## $ doors    : chr  "2" "2" "2" "2" ...
## $ persons  : chr  "2" "2" "2" "2" ...
## $ lug_boot : chr  "small" "small" "small" "med" ...
## $ safety   : chr  "low" "med" "high" "low" ...
## $ rating   : chr  "unacc" "unacc" "unacc" "unacc" ...
```

```
head(uciCar)
```

```
##   buying maint doors persons lug_boot safety rating
## 1  vhigh vhigh     2         2    small    low  unacc
## 2  vhigh vhigh     2         2    small    med  unacc
## 3  vhigh vhigh     2         2    small    high unacc
## 4  vhigh vhigh     2         2     med    low  unacc
## 5  vhigh vhigh     2         2     med    med  unacc
## 6  vhigh vhigh     2         2     med    high unacc
```

- Listing 2.3

```
d <- read.table(paste('http://archive.ics.uci.edu/ml/',
  'machine-learning-databases/statlog/german/german.data', sep=''),
  stringsAsFactors=F, header=F)
print(d[1:3,])
```

```
##      V1 V2  V3  V4   V5  V6  V7 V8  V9  V10 V11  V12 V13  V14  V15 V16  V17
## 1 A11  6 A34 A43 1169 A65 A75  4 A93 A101  4 A121  67 A143 A152  2 A173
## 2 A12 48 A32 A43 5951 A61 A73  2 A92 A101  2 A121  22 A143 A152  1 A173
## 3 A14 12 A34 A46 2096 A61 A74  2 A93 A101  3 A121  49 A143 A152  1 A172
##      V18  V19  V20 V21
## 1  1 A192 A201  1
## 2  1 A191 A201  2
## 3  2 A191 A201  1
```

- Listing 2.4

```
colnames(d) <- c('Status.of.existing.checking.account',
  'Duration.in.month', 'Credit.history', 'Purpose',
  'Credit.amount', 'Savings account/bonds',
  'Present.employment.since',
  'Installment.rate.in.percentage.of.disposable.income',
  'Personal.status.and.sex', 'Other.debtors/guarantors',
  'Present.residence.since', 'Property', 'Age.in.years',
  'Other.installment.plans', 'Housing',
  'Number.of.existing.credits.at.this.bank', 'Job',
  'Number.of.people.being.liable.to.provide.maintenance.for',
  'Telephone', 'foreign.worker', 'Good.Loan')
options(width=132)
print(d[1:3,])
```

```

##      Status.of.existing.checking.account Duration.in.month Credit.history Purpo
se Credit.amount Savings account/bonds
## 1              A11              6              A34      A
43          1169              A65
## 2              A12              48              A32      A
43          5951              A61
## 3              A14              12              A34      A
46          2096              A61
##      Present.employment.since Installment.rate.in.percentage.of.disposable.inco
me Personal.status.and.sex Other.debtors/guarantors
## 1              A75
4              A93              A101
## 2              A73
2              A92              A101
## 3              A74
2              A93              A101
##      Present.residence.since Property Age.in.years Other.installment.plans Hous
ing Number.of.existing.credits.at.this.bank Job
## 1              4      A121              67              A143      A
152              2 A173
## 2              2      A121              22              A143      A
152              1 A173
## 3              3      A121              49              A143      A
152              1 A172
##      Number.of.people.being.liable.to.provide.maintenance.for Telephone foreig
n.worker Good.Loan
## 1              1              A192
A201
## 2              1              A191
A201
## 3              2              A191
A201

```

- Listing 2.5

```

mapping <- list('A11'='... < 0 DM',
'A12'='0 <= ... < 200 DM',
'A13'='... >= 200 DM / salary assignments for at least 1 year',
'A14'='no checking account',
'A30'='no credits taken/all credits paid back duly',
'A31'='all credits at this bank paid back duly',
'A32'='existing credits paid back duly till now',
'A33'='delay in paying off in the past',
'A34'='critical account/other credits existing (not at this bank)',
'A40'='car (new)',
'A41'='car (used)',
'A42'='furniture/equipment',
'A43'='radio/television',
'A44'='domestic appliances',
'A45'='repairs',
'A46'='education',
'A47'='(vacation - does not exist?)',
'A48'='retraining',
'A49'='business',
'A410'='others',
'A61'='... < 100 DM',
'A62'='100 <= ... < 500 DM',
'A63'='500 <= ... < 1000 DM',
'A64'='.. >= 1000 DM',
'A65'='unknown/ no savings account',
'A71'='unemployed',
'A72'='... < 1 year',
'A73'='1 <= ... < 4 years',
'A74'='4 <= ... < 7 years',
'A75'='.. >= 7 years',
'A91'='male : divorced/separated',
'A92'='female : divorced/separated/married',
'A93'='male : single',
'A94'='male : married/widowed',
'A95'='female : single',
'A101'='none',
'A102'='co-applicant',
'A103'='guarantor',
'A121'='real estate',
'A122'='if not A121 : building society savings agreement/life insurance',
'A123'='if not A121/A122 : car or other, not in attribute 6',
'A124'='unknown / no property',
'A141'='bank',
'A142'='stores',
'A143'='none',
'A151'='rent',
'A152'='own',
'A153'='for free',
'A171'='unemployed/ unskilled - non-resident',
'A172'='unskilled - resident',
'A173'='skilled employee / official',
'A174'='management/ self-employed/highly qualified employee/ officer',
'A191'='none',

```

```
'A192'='yes, registered under the customers name',
'A201'='yes',
'A202'='no')
```

- Listing 2.6의 구조를 파악하기 위해서

```
sapply(d, class)
```

```
##              Status.of.existing.checking.account
Duration.in.month
##              "character"
"integer"
##              Credit.history
Purpose
##              "character"
"character"
##              Credit.amount
Savings account/bonds
##              "integer"
"character"
##              Present.employment.since      Installment.ra
te.in.percentage.of.disposable.income
##              "character"
"integer"
##              Personal.status.and.sex
Other.debtors/guarantors
##              "character"
"character"
##              Present.residence.since
Property
##              "integer"
"character"
##              Age.in.years
Other.installment.plans
##              "integer"
"character"
##              Housing                        Nu
mber.of.existing.credits.at.this.bank
##              "character"
"integer"
##              Job Number.of.people.be
ing.liable.to.provide.maintenance.for
##              "character"
"integer"
##              Telephone
foreign.worker
##              "character"
"character"
##              Good.Loan
##              "integer"
```

```
head(d[, 4])
```

```
## [1] "A43" "A43" "A46" "A42" "A40" "A46"
```

```
mapping[head(d[, 4])]
```

```
## $A43
## [1] "radio/television"
##
## $A43
## [1] "radio/television"
##
## $A46
## [1] "education"
##
## $A42
## [1] "furniture/equipment"
##
## $A40
## [1] "car (new)"
##
## $A46
## [1] "education"
```

```
as.character(mapping[head(d[, 4])])
```

```
## [1] "radio/television"      "radio/television"      "education"              "furni
ture/equipment" "car (new)"
## [6] "education"
```

- d 의 복제를 만들어 두고,

```
d.2 <- d
```

- Listing 2.6

```
for(i in 1:dim(d)[2]) {
  if(class(d[, i])=="character") {
    d[, i] <- as.factor(as.character(mapping[d[, i]]))
  }
}
str(d)
```

```
## 'data.frame':    1000 obs. of  21 variables:
## $ Status.of.existing.checking.account      : Factor w/ 4 lev
els "... < 0 DM", "... >= 200 DM / salary assignments for at least 1 year",...: 1
3 4 1 1 4 4 3 4 3 ...
## $ Duration.in.month                       : int   6 48 12 42
24 36 24 36 12 30 ...
## $ Credit.history                          : Factor w/ 5 lev
els "all credits at this bank paid back duly",...: 2 4 2 4 3 4 4 4 4 2 ...
## $ Purpose                                : Factor w/ 10 le
vels "business", "car (new)",...: 8 8 5 6 2 5 6 3 8 2 ...
## $ Credit.amount                          : int  1169 5951
2096 7882 4870 9055 2835 6948 3059 5234 ...
## $ Savings account/bonds                  : Factor w/ 5 lev
els "... >= 1000 DM",...: 5 2 2 2 2 5 4 2 1 2 ...
## $ Present.employment.since               : Factor w/ 5 lev
els "... >= 7 years",...: 1 3 4 4 3 3 1 3 4 5 ...
## $ Installment.rate.in.percentage.of.disposable.income : int   4 2 2 2 3
2 3 2 2 4 ...
## $ Personal.status.and.sex                : Factor w/ 4 lev
els "female : divorced/separated/married",...: 4 1 4 4 4 4 4 4 2 3 ...
## $ Other.debtors/guarantors               : Factor w/ 3 lev
els "co-applicant",...: 3 3 3 2 3 3 3 3 3 3 ...
## $ Present.residence.since                : int   4 2 3 4 4
4 4 2 4 2 ...
## $ Property                              : Factor w/ 4 lev
els "if not A121 : building society savings agreement/life insurance",...: 3 3 3
1 4 4 1 2 3 2 ...
## $ Age.in.years                           : int   67 22 49 4
5 53 35 53 35 61 28 ...
## $ Other.installment.plans                : Factor w/ 3 lev
els "bank", "none",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ Housing                               : Factor w/ 3 lev
els "for free", "own",...: 2 2 2 1 1 1 2 3 2 2 ...
## $ Number.of.existing.credits.at.this.bank : int   2 1 1 1 2
1 1 1 1 2 ...
## $ Job                                    : Factor w/ 4 lev
els "management/ self-employed/highly qualified employee/ officer",...: 2 2 4 2
2 4 2 1 4 1 ...
## $ Number.of.people.being.liable.to.provide.maintenance.for: int   1 1 2 2 2
2 1 1 1 1 ...
## $ Telephone                             : Factor w/ 2 lev
els "none", "yes, registered under the customers name": 2 1 1 1 1 2 1 2 1 1 ...
## $ foreign.worker                         : Factor w/ 2 lev
els "no", "yes": 2 2 2 2 2 2 2 2 2 2 ...
## $ Good.Loan                             : int   1 2 1 1 2
1 1 1 1 2 ...
```

```
d$Good.Loan <- factor(ifelse(d$Good.Loan == 1, "GoodLoan", "BadLoan"))
str(d)
```

```
## 'data.frame':    1000 obs. of  21 variables:
## $ Status.of.existing.checking.account      : Factor w/ 4 lev
els "... < 0 DM", "... >= 200 DM / salary assignments for at least 1 year",...: 1
3 4 1 1 4 4 3 4 3 ...
## $ Duration.in.month                       : int   6 48 12 42
24 36 24 36 12 30 ...
## $ Credit.history                          : Factor w/ 5 lev
els "all credits at this bank paid back duly",...: 2 4 2 4 3 4 4 4 4 2 ...
## $ Purpose                                : Factor w/ 10 le
vels "business", "car (new)",...: 8 8 5 6 2 5 6 3 8 2 ...
## $ Credit.amount                          : int  1169 5951
2096 7882 4870 9055 2835 6948 3059 5234 ...
## $ Savings account/bonds                  : Factor w/ 5 lev
els "... >= 1000 DM",...: 5 2 2 2 2 5 4 2 1 2 ...
## $ Present.employment.since               : Factor w/ 5 lev
els "... >= 7 years",...: 1 3 4 4 3 3 1 3 4 5 ...
## $ Installment.rate.in.percentage.of.disposable.income : int   4 2 2 2 3
2 3 2 2 4 ...
## $ Personal.status.and.sex                : Factor w/ 4 lev
els "female : divorced/separated/married",...: 4 1 4 4 4 4 4 4 2 3 ...
## $ Other.debtors/guarantors               : Factor w/ 3 lev
els "co-applicant",...: 3 3 3 2 3 3 3 3 3 3 ...
## $ Present.residence.since                : int   4 2 3 4 4
4 4 2 4 2 ...
## $ Property                              : Factor w/ 4 lev
els "if not A121 : building society savings agreement/life insurance",...: 3 3 3
1 4 4 1 2 3 2 ...
## $ Age.in.years                          : int   67 22 49 4
5 53 35 53 35 61 28 ...
## $ Other.installment.plans                : Factor w/ 3 lev
els "bank", "none",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ Housing                              : Factor w/ 3 lev
els "for free", "own",...: 2 2 2 1 1 1 2 3 2 2 ...
## $ Number.of.existing.credits.at.this.bank : int   2 1 1 1 2
1 1 1 1 2 ...
## $ Job                                  : Factor w/ 4 lev
els "management/ self-employed/highly qualified employee/ officer",...: 2 2 4 2
2 4 2 1 4 1 ...
## $ Number.of.people.being.liable.to.provide.maintenance.for: int   1 1 2 2 2
2 1 1 1 1 ...
## $ Telephone                            : Factor w/ 2 lev
els "none", "yes, registered under the customers name": 2 1 1 1 1 2 1 2 1 1 ...
## $ foreign.worker                        : Factor w/ 2 lev
els "no", "yes": 2 2 2 2 2 2 2 2 2 2 ...
## $ Good.Loan                            : Factor w/ 2 lev
els "BadLoan", "GoodLoan": 2 1 2 2 1 2 2 2 2 1 ...
```

- for loop이 보기 싫다면,

```
l.char <- sapply(d.2, class) == "character"
l.char
```



```

##                Status.of.existing.checking.account
Duration.in.month
##                                TRUE
FALSE
##                Credit.history
Purpose
##                                TRUE
TRUE
##                Credit.amount
Savings account/bonds
##                                FALSE
TRUE
##                Present.employment.since      Installment.ra
te.in.percentage.of.disposable.income
##                                TRUE
FALSE
##                Personal.status.and.sex
Other.debtors/guarantors
##                                TRUE
TRUE
##                Present.residence.since
Property
##                                FALSE
TRUE
##                Age.in.years
Other.installment.plans
##                                FALSE
TRUE
##                Housing                        Nu
mber.of.existing.credits.at.this.bank
##                                TRUE
FALSE
##                Job Number.of.people.be
ing.liable.to.provide.maintenance.for
##                                TRUE
FALSE
##                Telephone
foreign.worker
##                                TRUE
TRUE
##                Good.Loan
##                                FALSE

```

```

names(l.char) <- NULL
l.char

```

```

## [1] TRUE FALSE TRUE TRUE FALSE TRUE TRUE FALSE TRUE TRUE FALSE TRUE
FALSE TRUE TRUE FALSE TRUE FALSE TRUE TRUE FALSE

```

```

which(l.char)

```

```
## [1] 1 3 4 6 7 9 10 12 14 15 17 19 20
```

```
d.char <- subset(d.2, select = l.char)
str(d.char)
```

```
## 'data.frame': 1000 obs. of 13 variables:
## $ Status.of.existing.checking.account: chr "A11" "A12" "A14" "A11" ...
## $ Credit.history : chr "A34" "A32" "A34" "A32" ...
## $ Purpose : chr "A43" "A43" "A46" "A42" ...
## $ Savings account/bonds : chr "A65" "A61" "A61" "A61" ...
## $ Present.employment.since : chr "A75" "A73" "A74" "A74" ...
## $ Personal.status.and.sex : chr "A93" "A92" "A93" "A93" ...
## $ Other.debtors/guarantors : chr "A101" "A101" "A101" "A103" ...
## $ Property : chr "A121" "A121" "A121" "A122" ...
## $ Other.installment.plans : chr "A143" "A143" "A143" "A143" ...
## $ Housing : chr "A152" "A152" "A152" "A153" ...
## $ Job : chr "A173" "A173" "A172" "A173" ...
## $ Telephone : chr "A192" "A191" "A191" "A191" ...
## $ foreign.worker : chr "A201" "A201" "A201" "A201" ...
```

```
d.factor <- sapply(d.char, function(x) factor(as.character(mapping[x])), simpli
fy=FALSE)
str(d.factor)
```

```
## List of 13
## $ Status.of.existing.checking.account: Factor w/ 4 levels "... < 0 DM","...
>= 200 DM / salary assignments for at least 1 year",...: 1 3 4 1 1 4 4 3 4 3 ...
## $ Credit.history : Factor w/ 5 levels "all credits at t
his bank paid back duly",...: 2 4 2 4 3 4 4 4 4 2 ...
## $ Purpose : Factor w/ 10 levels "business","car
(new)",...: 8 8 5 6 2 5 6 3 8 2 ...
## $ Savings account/bonds : Factor w/ 5 levels ".. >= 1000 D
M",...: 5 2 2 2 2 5 4 2 1 2 ...
## $ Present.employment.since : Factor w/ 5 levels ".. >= 7 year
s",...: 1 3 4 4 3 3 1 3 4 5 ...
## $ Personal.status.and.sex : Factor w/ 4 levels "female : divorce
d/separated/married",...: 4 1 4 4 4 4 4 4 2 3 ...
## $ Other.debtors/guarantors : Factor w/ 3 levels "co-applican
t",...: 3 3 3 2 3 3 3 3 3 3 ...
## $ Property : Factor w/ 4 levels "if not A121 : bu
ilding society savings agreement/life insurance",...: 3 3 3 1 4 4 1 2 3 2 ...
## $ Other.installment.plans : Factor w/ 3 levels "bank","none",...:
2 2 2 2 2 2 2 2 2 2 ...
## $ Housing : Factor w/ 3 levels "for free","ow
n",...: 2 2 2 1 1 1 2 3 2 2 ...
## $ Job : Factor w/ 4 levels "management/ sel
f-employed/highly qualified employee/ officer",...: 2 2 4 2 2 4 2 1 4 1 ...
## $ Telephone : Factor w/ 2 levels "none","yes, regi
stered under the customers name": 2 1 1 1 1 2 1 2 1 1 ...
## $ foreign.worker : Factor w/ 2 levels "no","yes": 2 2 2
2 2 2 2 2 2 2 ...
```

```
d.2[l.char] <- d.factor
str(d.2)
```

```
## 'data.frame':    1000 obs. of  21 variables:
## $ Status.of.existing.checking.account      : Factor w/ 4 lev
els "... < 0 DM", "... >= 200 DM / salary assignments for at least 1 year",...: 1
3 4 1 1 4 4 3 4 3 ...
## $ Duration.in.month                      : int   6 48 12 42
24 36 24 36 12 30 ...
## $ Credit.history                        : Factor w/ 5 lev
els "all credits at this bank paid back duly",...: 2 4 2 4 3 4 4 4 4 2 ...
## $ Purpose                              : Factor w/ 10 le
vels "business", "car (new)",...: 8 8 5 6 2 5 6 3 8 2 ...
## $ Credit.amount                        : int  1169 5951
2096 7882 4870 9055 2835 6948 3059 5234 ...
## $ Savings account/bonds                : Factor w/ 5 lev
els "... >= 1000 DM",...: 5 2 2 2 2 5 4 2 1 2 ...
## $ Present.employment.since              : Factor w/ 5 lev
els "... >= 7 years",...: 1 3 4 4 3 3 1 3 4 5 ...
## $ Installment.rate.in.percentage.of.disposable.income : int   4 2 2 2 3
2 3 2 2 4 ...
## $ Personal.status.and.sex               : Factor w/ 4 lev
els "female : divorced/separated/married",...: 4 1 4 4 4 4 4 4 2 3 ...
## $ Other.debtors/guarantors              : Factor w/ 3 lev
els "co-applicant",...: 3 3 3 2 3 3 3 3 3 3 ...
## $ Present.residence.since               : int   4 2 3 4 4
4 4 2 4 2 ...
## $ Property                             : Factor w/ 4 lev
els "if not A121 : building society savings agreement/life insurance",...: 3 3 3
1 4 4 1 2 3 2 ...
## $ Age.in.years                         : int   67 22 49 4
5 53 35 53 35 61 28 ...
## $ Other.installment.plans               : Factor w/ 3 lev
els "bank", "none",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ Housing                             : Factor w/ 3 lev
els "for free", "own",...: 2 2 2 1 1 1 2 3 2 2 ...
## $ Number.of.existing.credits.at.this.bank : int   2 1 1 1 2
1 1 1 1 2 ...
## $ Job                                  : Factor w/ 4 lev
els "management/ self-employed/highly qualified employee/ officer",...: 2 2 4 2
2 4 2 1 4 1 ...
## $ Number.of.people.being.liable.to.provide.maintenance.for: int   1 1 2 2 2
2 1 1 1 1 ...
## $ Telephone                           : Factor w/ 2 lev
els "none", "yes, registered under the customers name": 2 1 1 1 1 2 1 2 1 1 ...
## $ foreign.worker                       : Factor w/ 2 lev
els "no", "yes": 2 2 2 2 2 2 2 2 2 2 ...
## $ Good.Loan                           : int   1 2 1 1 2
1 1 1 1 2 ...
```

```
d.2$Good.Loan <- factor(ifelse(d.2$Good.Loan == 1, "GoodLoan", "BadLoan"))
str(d.2)
```

```
## 'data.frame':    1000 obs. of  21 variables:
##  $ Status.of.existing.checking.account      : Factor w/ 4 lev
els "... < 0 DM","... >= 200 DM / salary assignments for at least 1 year",...: 1
3 4 1 1 4 4 3 4 3 ...
##  $ Duration.in.month                      : int   6 48 12 42
24 36 24 36 12 30 ...
##  $ Credit.history                        : Factor w/ 5 lev
els "all credits at this bank paid back duly",...: 2 4 2 4 3 4 4 4 4 2 ...
##  $ Purpose                              : Factor w/ 10 le
vels "business","car (new)",...: 8 8 5 6 2 5 6 3 8 2 ...
##  $ Credit.amount                        : int  1169 5951
2096 7882 4870 9055 2835 6948 3059 5234 ...
##  $ Savings account/bonds                : Factor w/ 5 lev
els "... >= 1000 DM",...: 5 2 2 2 2 5 4 2 1 2 ...
##  $ Present.employment.since              : Factor w/ 5 lev
els "... >= 7 years",...: 1 3 4 4 3 3 1 3 4 5 ...
##  $ Installment.rate.in.percentage.of.disposable.income : int   4 2 2 2 3
2 3 2 2 4 ...
##  $ Personal.status.and.sex               : Factor w/ 4 lev
els "female : divorced/separated/married",...: 4 1 4 4 4 4 4 4 2 3 ...
##  $ Other.debtors/guarantors              : Factor w/ 3 lev
els "co-applicant",...: 3 3 3 2 3 3 3 3 3 3 ...
##  $ Present.residence.since               : int   4 2 3 4 4
4 4 2 4 2 ...
##  $ Property                             : Factor w/ 4 lev
els "if not A121 : building society savings agreement/life insurance",...: 3 3 3
1 4 4 1 2 3 2 ...
##  $ Age.in.years                         : int   67 22 49 4
5 53 35 53 35 61 28 ...
##  $ Other.installment.plans               : Factor w/ 3 lev
els "bank","none",...: 2 2 2 2 2 2 2 2 2 2 ...
##  $ Housing                             : Factor w/ 3 lev
els "for free","own",...: 2 2 2 1 1 1 2 3 2 2 ...
##  $ Number.of.existing.credits.at.this.bank : int   2 1 1 1 2
1 1 1 1 2 ...
##  $ Job                                  : Factor w/ 4 lev
els "management/ self-employed/highly qualified employee/ officer",...: 2 2 4 2
2 4 2 1 4 1 ...
##  $ Number.of.people.being.liable.to.provide.maintenance.for: int   1 1 2 2 2
2 1 1 1 1 ...
##  $ Telephone                           : Factor w/ 2 lev
els "none","yes, registered under the customers name": 2 1 1 1 1 2 1 2 1 1 ...
##  $ foreign.worker                       : Factor w/ 2 lev
els "no","yes": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Good.Loan                           : Factor w/ 2 lev
els "BadLoan","GoodLoan": 2 1 2 2 1 2 2 2 2 1 ...
```

- 몇 가지 자료 검색

```
d[1:3, "Purpose"]
```

```
## [1] radio/television radio/television education
## 10 Levels: business car (new) car (used) domestic appliances education furni
ture/equipment others radio/television ... retraining
```

```
summary(d$Purpose)
```

```
##          business          car (new)          car (used) domestic applian
ces          education furniture/equipment
##          97          234          103
12          50          181
##          others    radio/television          repairs          retrain
ing
##          12          280          22
9
```

- Listing 7

```
(tbl <- table("Purpose" = d$Purpose, "Good Loan?" = d$Good.Loan))
```

```
##          Good Loan?
## Purpose          BadLoan GoodLoan
## business          34          63
## car (new)          89          145
## car (used)         17          86
## domestic appliances  4           8
## education          22          28
## furniture/equipment 58          123
## others              5           7
## radio/television    62          218
## repairs             8           14
## retraining          1           8
```

```
(tbl.df <- data.frame(tbl))
```

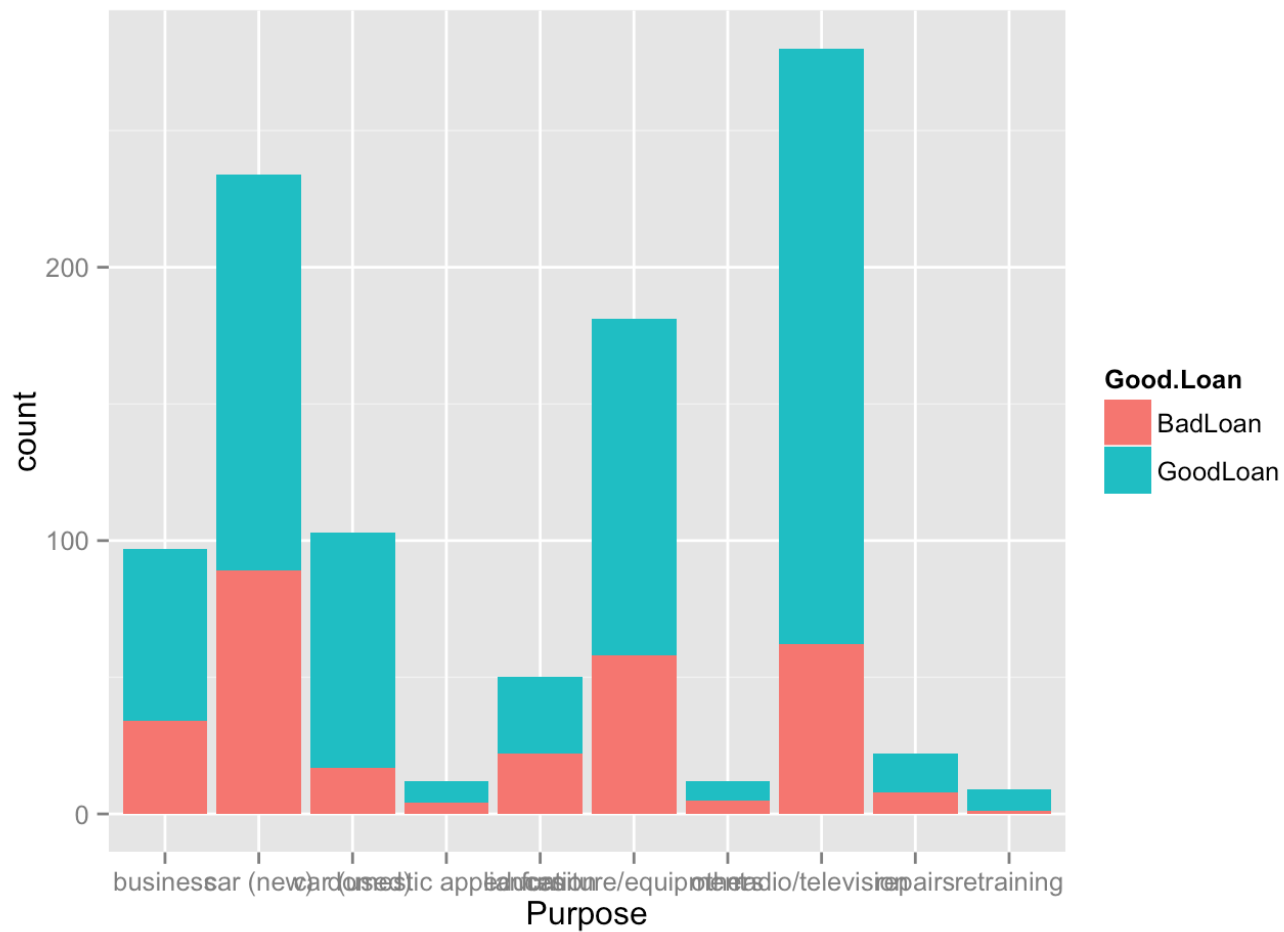
```
##           Purpose Good.Loan. Freq
## 1      business    BadLoan   34
## 2      car (new)    BadLoan   89
## 3      car (used)   BadLoan   17
## 4 domestic appliances BadLoan    4
## 5      education    BadLoan   22
## 6 furniture/equipment BadLoan   58
## 7      others       BadLoan    5
## 8      radio/television BadLoan  62
## 9      repairs      BadLoan    8
## 10     retraining   BadLoan    1
## 11     business     GoodLoan   63
## 12     car (new)     GoodLoan  145
## 13     car (used)    GoodLoan   86
## 14 domestic appliances GoodLoan    8
## 15     education    GoodLoan   28
## 16 furniture/equipment GoodLoan  123
## 17     others       GoodLoan    7
## 18     radio/television GoodLoan 218
## 19     repairs      GoodLoan   14
## 20     retraining   GoodLoan    8
```

```
options(digits=2)
prop.table(table("Purpose" = d$Purpose, "Good Loan?" = d$Good.Loan), 1)
```

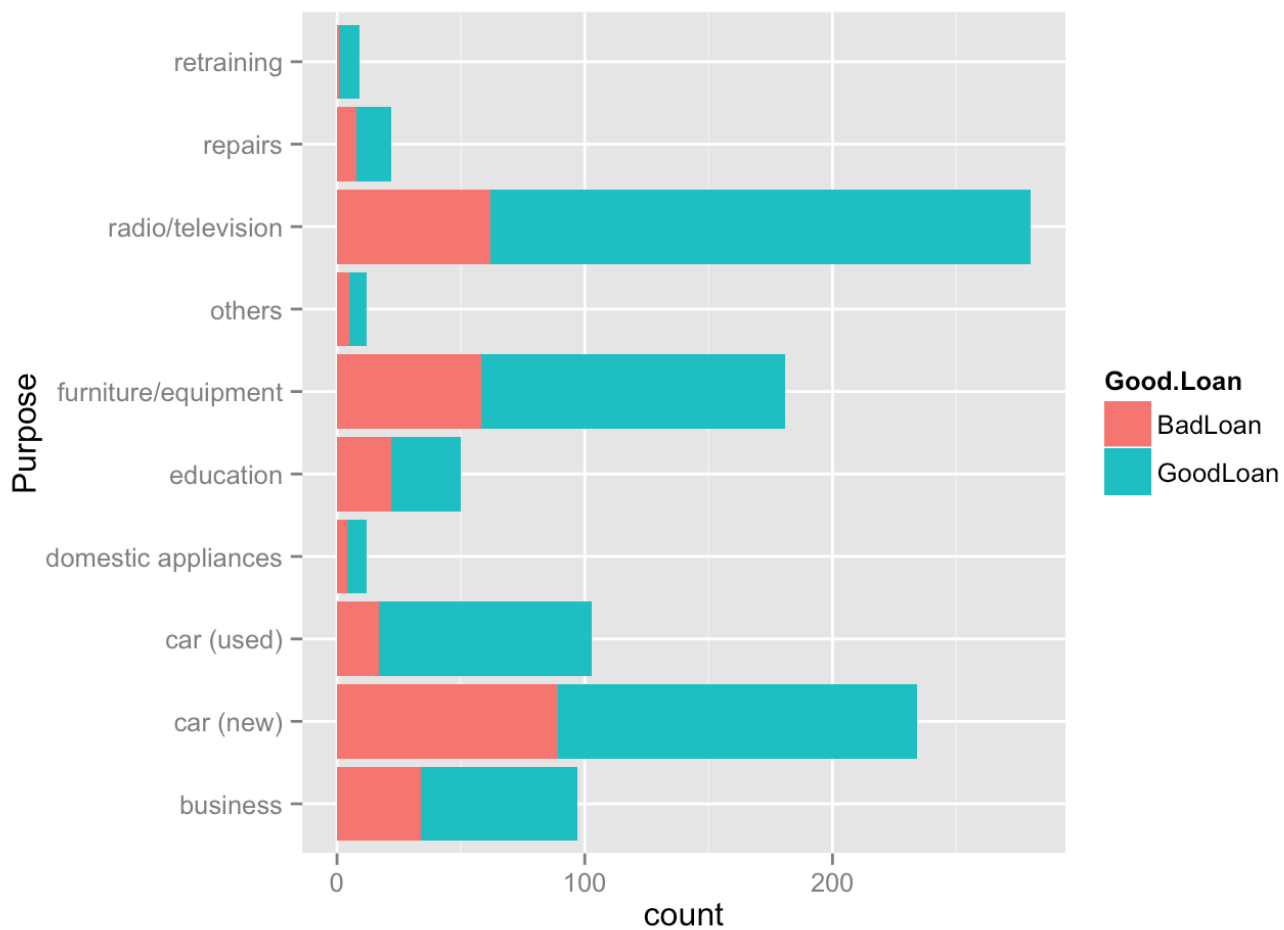
```
##           Good Loan?
## Purpose      BadLoan GoodLoan
## business      0.35    0.65
## car (new)      0.38    0.62
## car (used)     0.17    0.83
## domestic appliances 0.33    0.67
## education      0.44    0.56
## furniture/equipment 0.32    0.68
## others         0.42    0.58
## radio/television 0.22    0.78
## repairs        0.36    0.64
## retraining     0.11    0.89
```

- ggplot

```
library(ggplot2)
ggplot(d) + geom_bar(aes(x=Purpose, fill=Good.Loan))
```

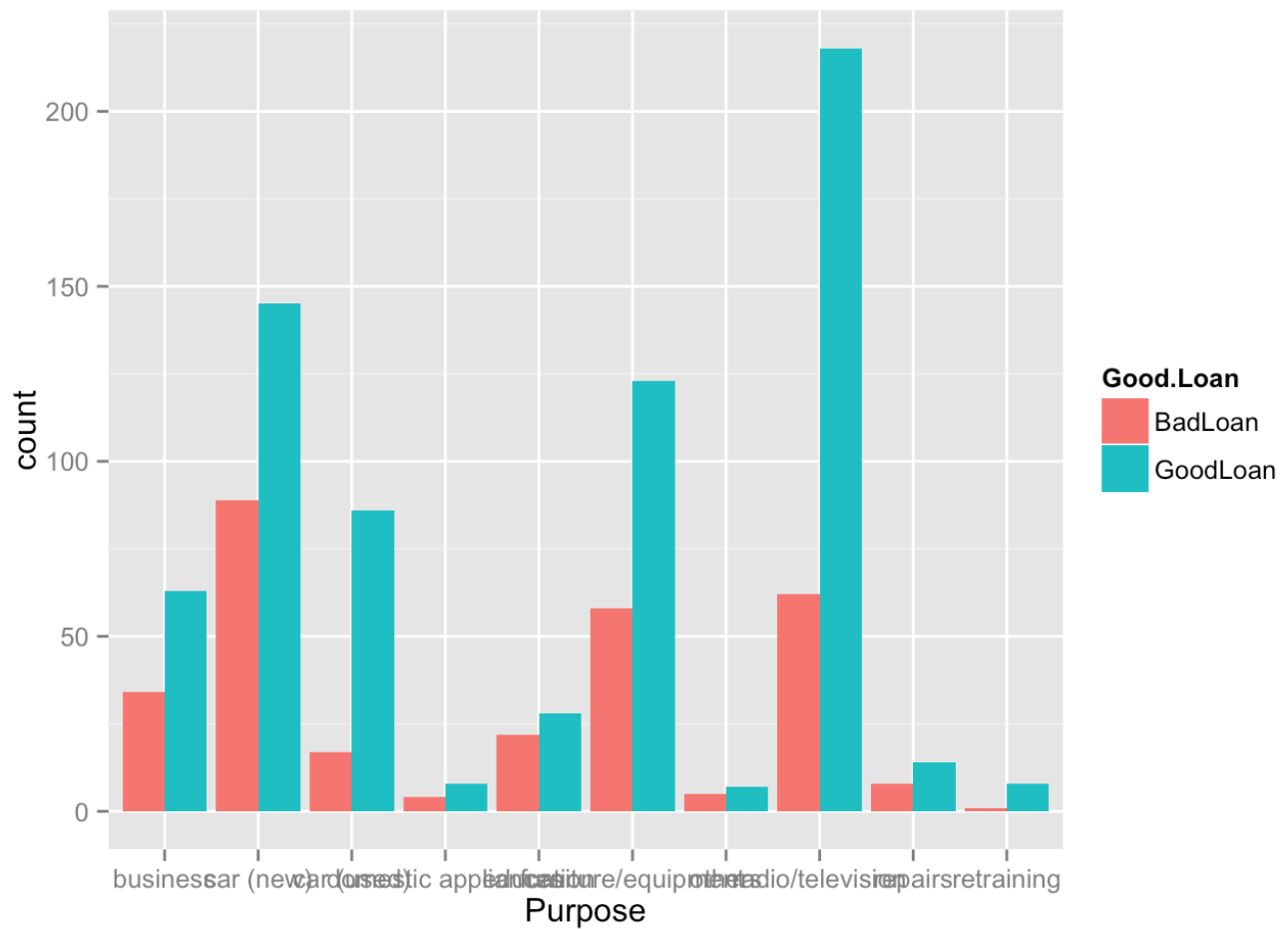


```
ggplot(d) + geom_bar(aes(x=Purpose, fill=Good.Loan)) + coord_flip()
```

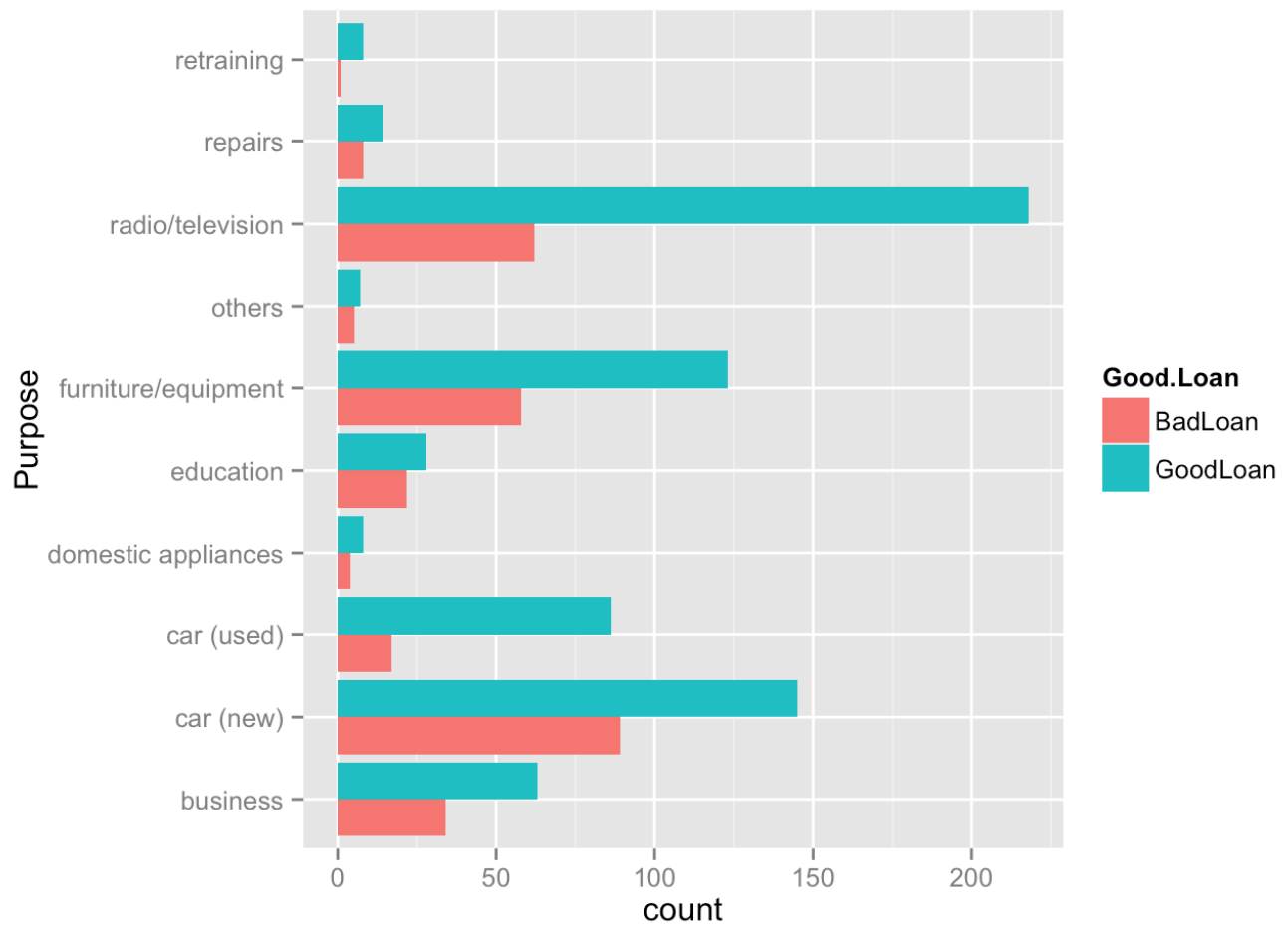




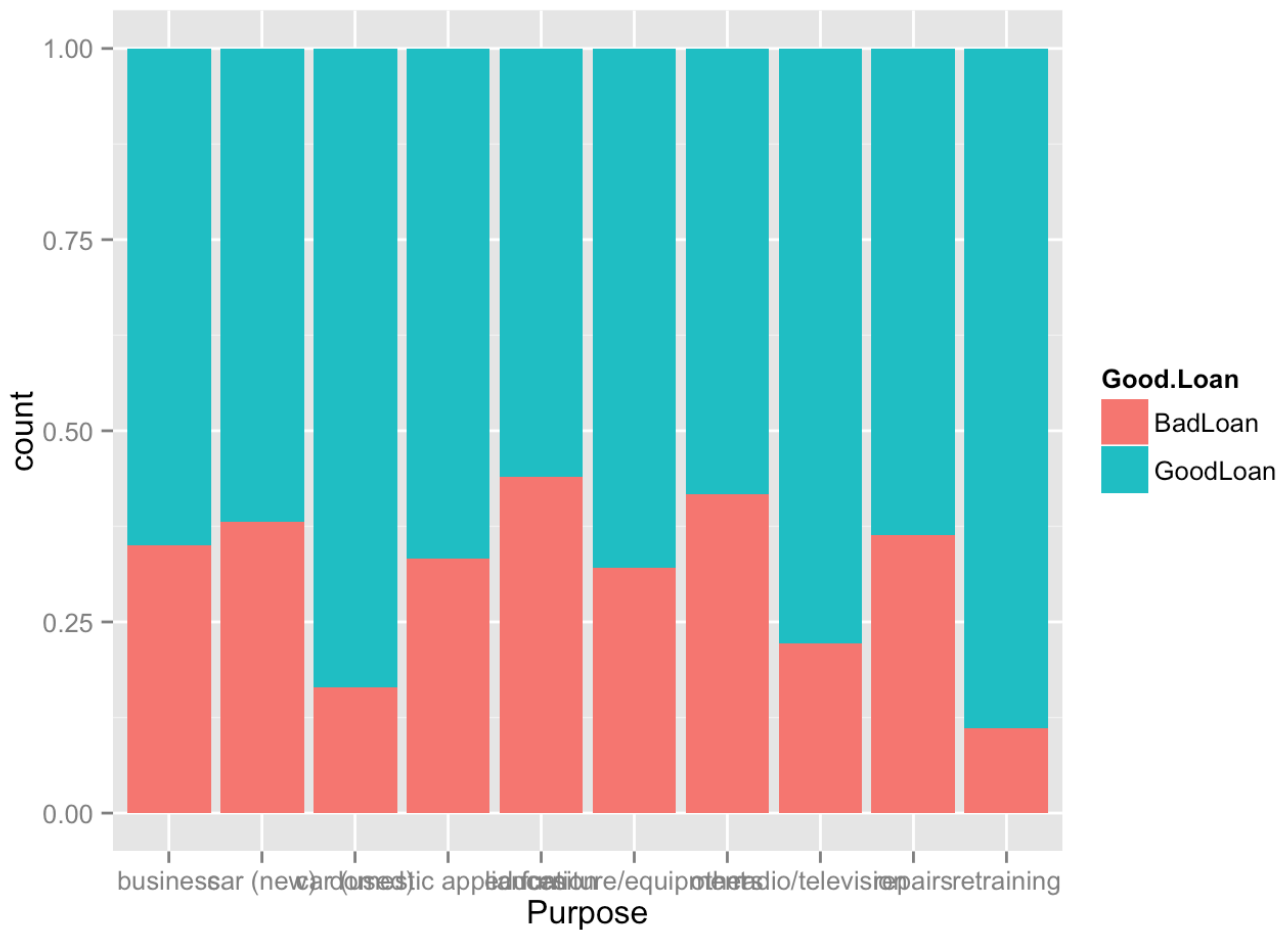
```
ggplot(d) + geom_bar(aes(x=Purpose, fill=Good.Loan), position="dodge")
```



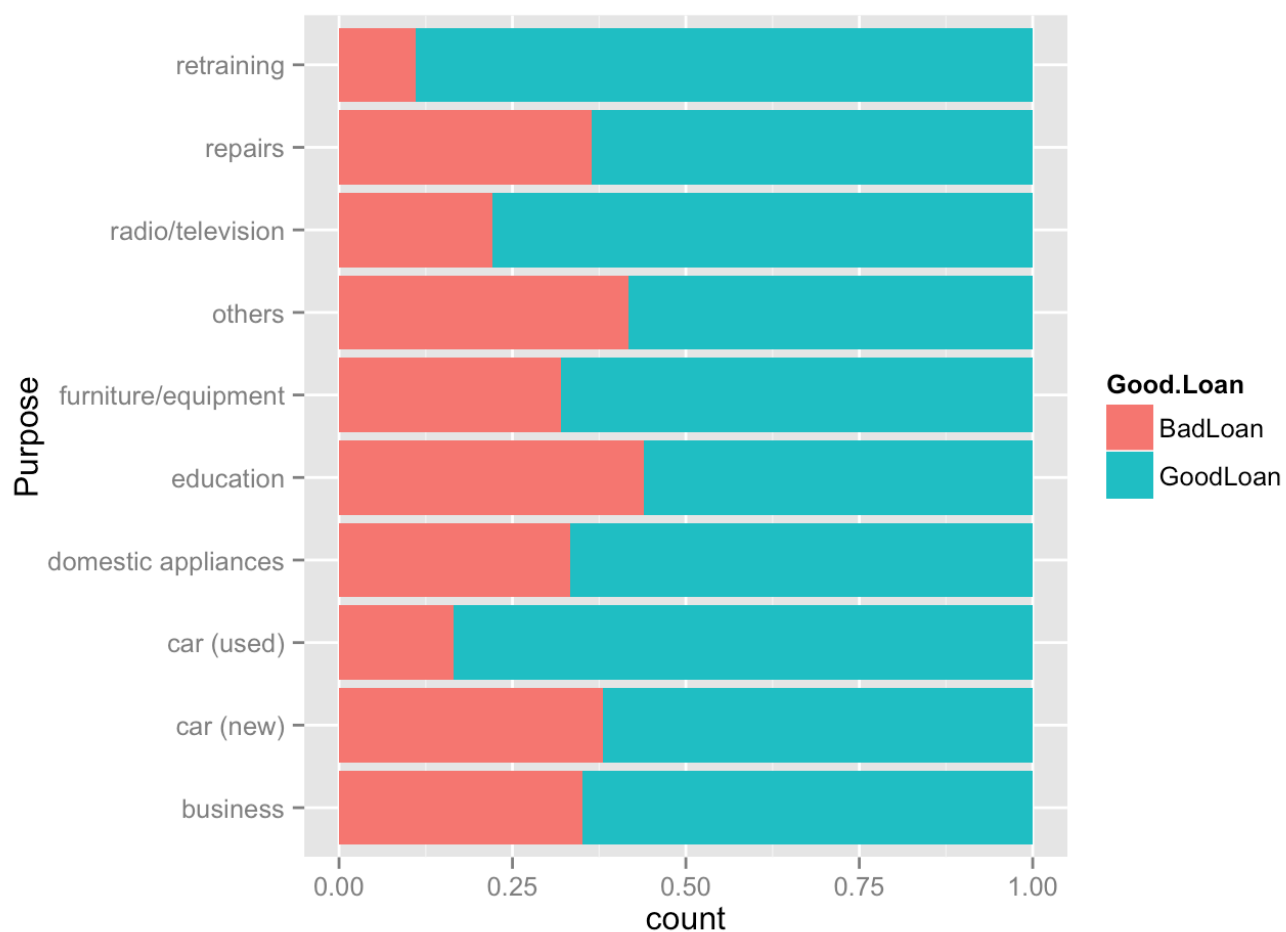
```
ggplot(d) + geom_bar(aes(x=Purpose, fill=Good.Loan), position="dodge") + coord_flip()
```



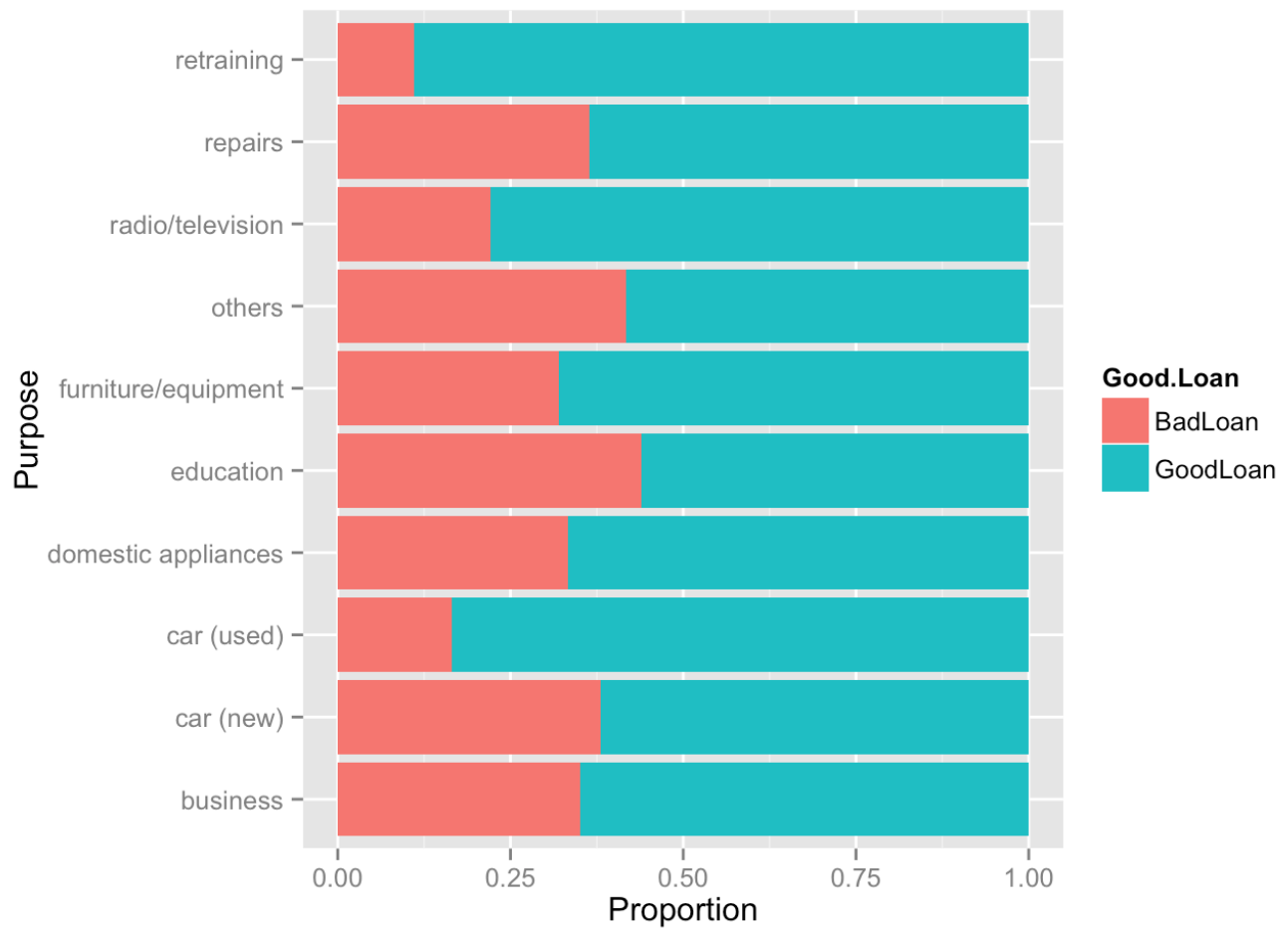
```
ggplot(d) + geom_bar(aes(x=Purpose, fill=Good.Loan), position="fill")
```



```
ggplot(d) + geom_bar(aes(x=Purpose, fill=Good.Loan), position="fill") + coord_f  
lip()
```



```
ggplot(d) + geom_bar(aes(x=Purpose, fill=Good.Loan), position="fill") + coord_f  
lip() + ylab("Proportion")
```

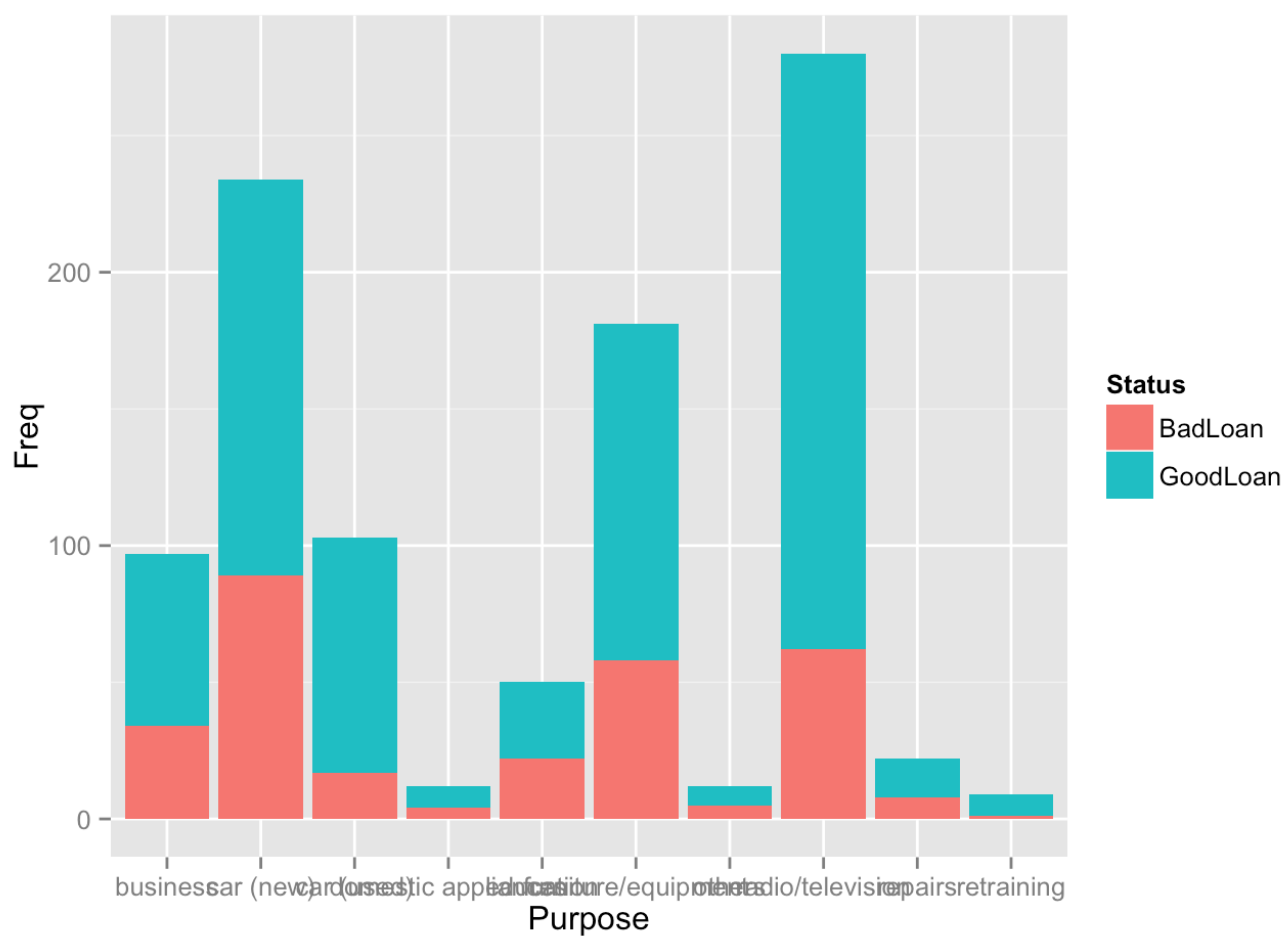


- `tbl.df` 로 그리면,

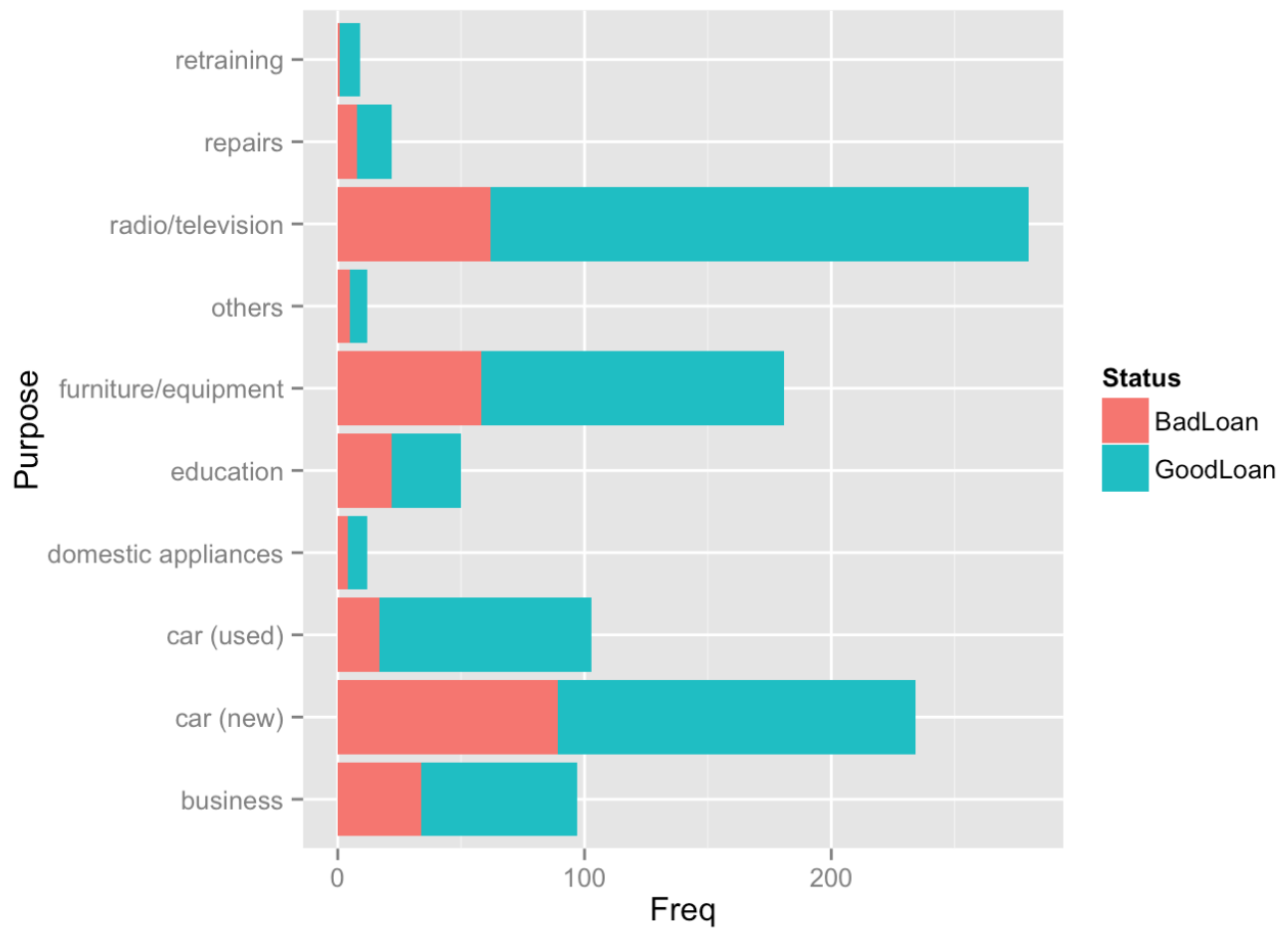
`tbl.df`

```
##           Purpose Good.Loan. Freq
## 1      business   BadLoan    34
## 2    car (new)   BadLoan    89
## 3    car (used)   BadLoan    17
## 4 domestic appliances BadLoan     4
## 5      education   BadLoan    22
## 6 furniture/equipment BadLoan    58
## 7         others   BadLoan     5
## 8  radio/television   BadLoan    62
## 9         repairs   BadLoan     8
## 10    retraining   BadLoan     1
## 11      business   GoodLoan    63
## 12    car (new)   GoodLoan   145
## 13    car (used)   GoodLoan    86
## 14 domestic appliances GoodLoan     8
## 15      education   GoodLoan    28
## 16 furniture/equipment GoodLoan   123
## 17         others   GoodLoan     7
## 18  radio/television   GoodLoan   218
## 19         repairs   GoodLoan    14
## 20    retraining   GoodLoan     8
```

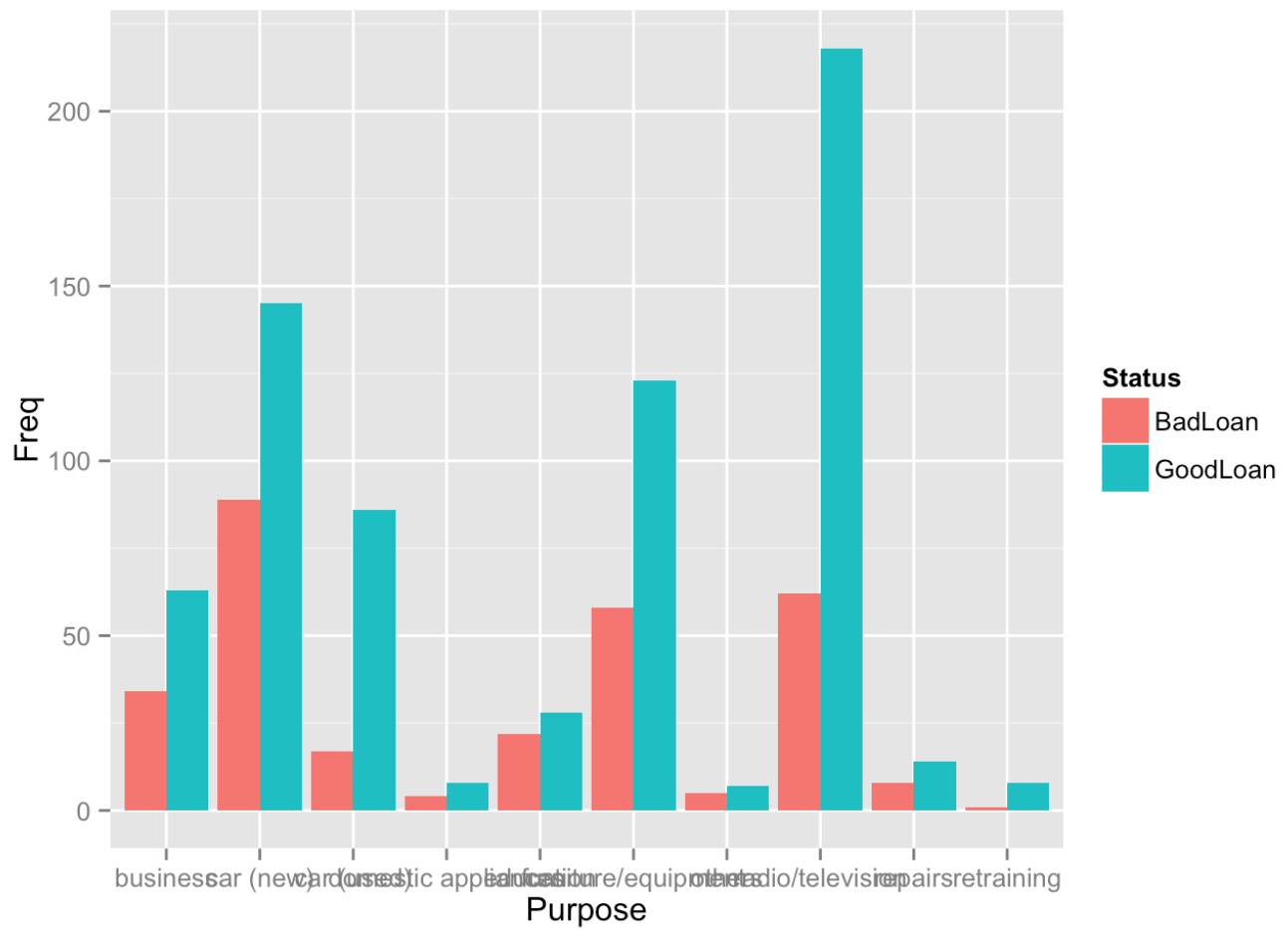
```
names(tbl.df)[2] <- c("Status")
ggplot(tbl.df, aes(x=Purpose, y=Freq, fill=Status)) + geom_bar(stat="identity")
```



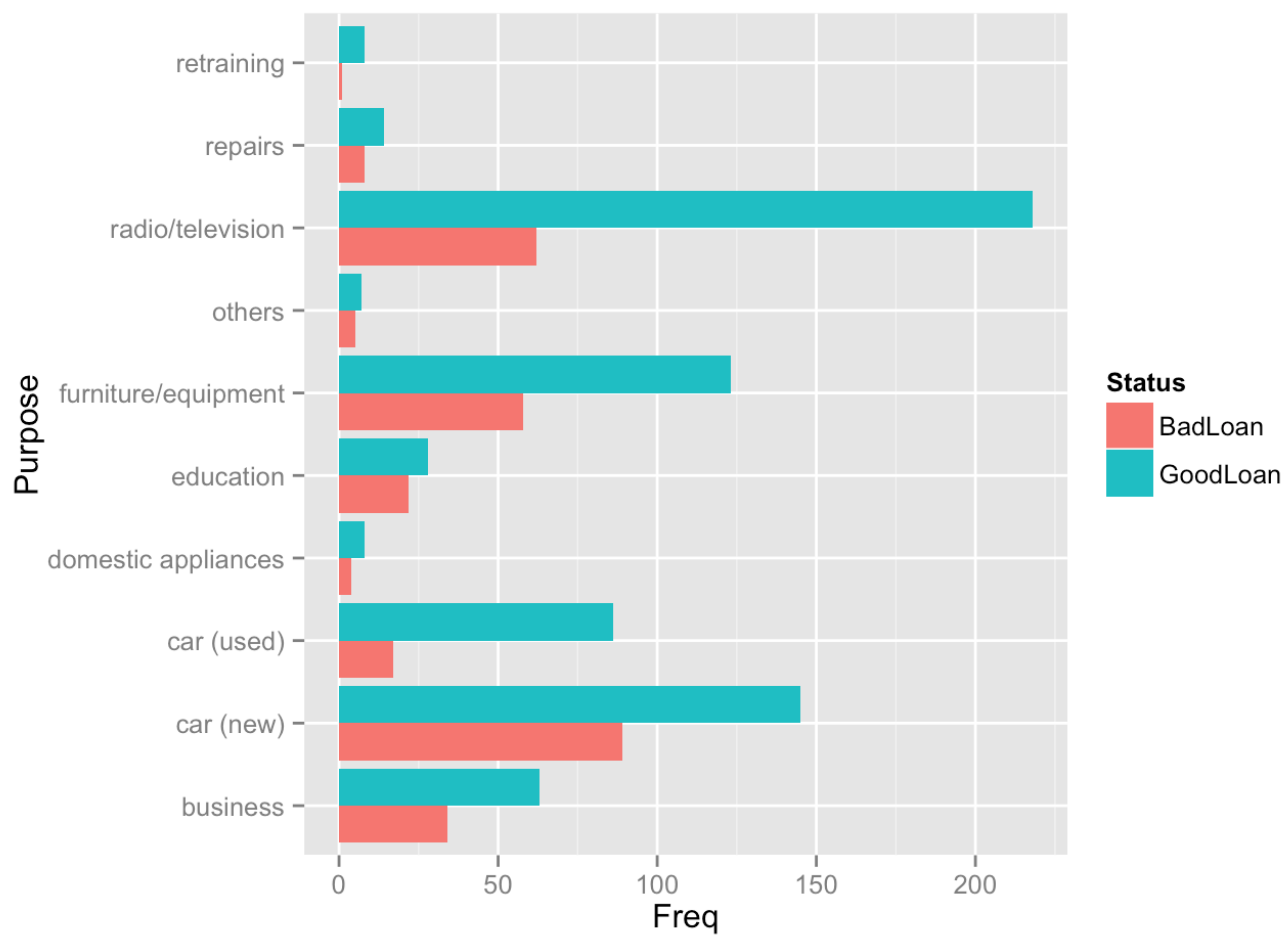
```
ggplot(tbl.df, aes(x=Purpose, y=Freq, fill=Status)) + geom_bar(stat="identity")
+
  coord_flip()
```



```
ggplot(tbl.df, aes(x=Purpose, y=Freq, fill=Status)) + geom_bar(stat="identity",  
position="dodge")
```

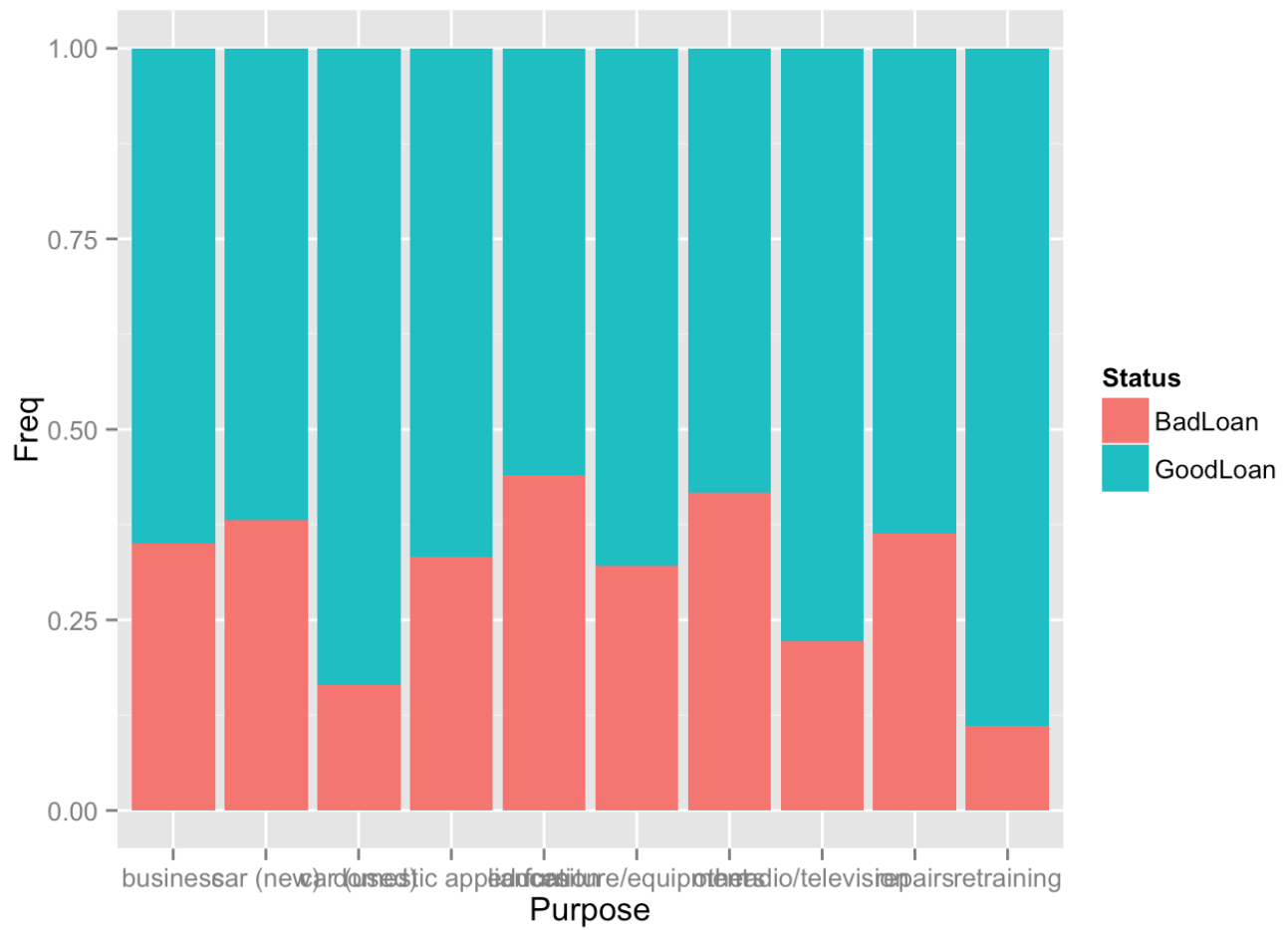


```
ggplot(tbl.df, aes(x=Purpose, y=Freq, fill=Status)) + geom_bar(stat="identity",
position="dodge") +
  coord_flip()
```

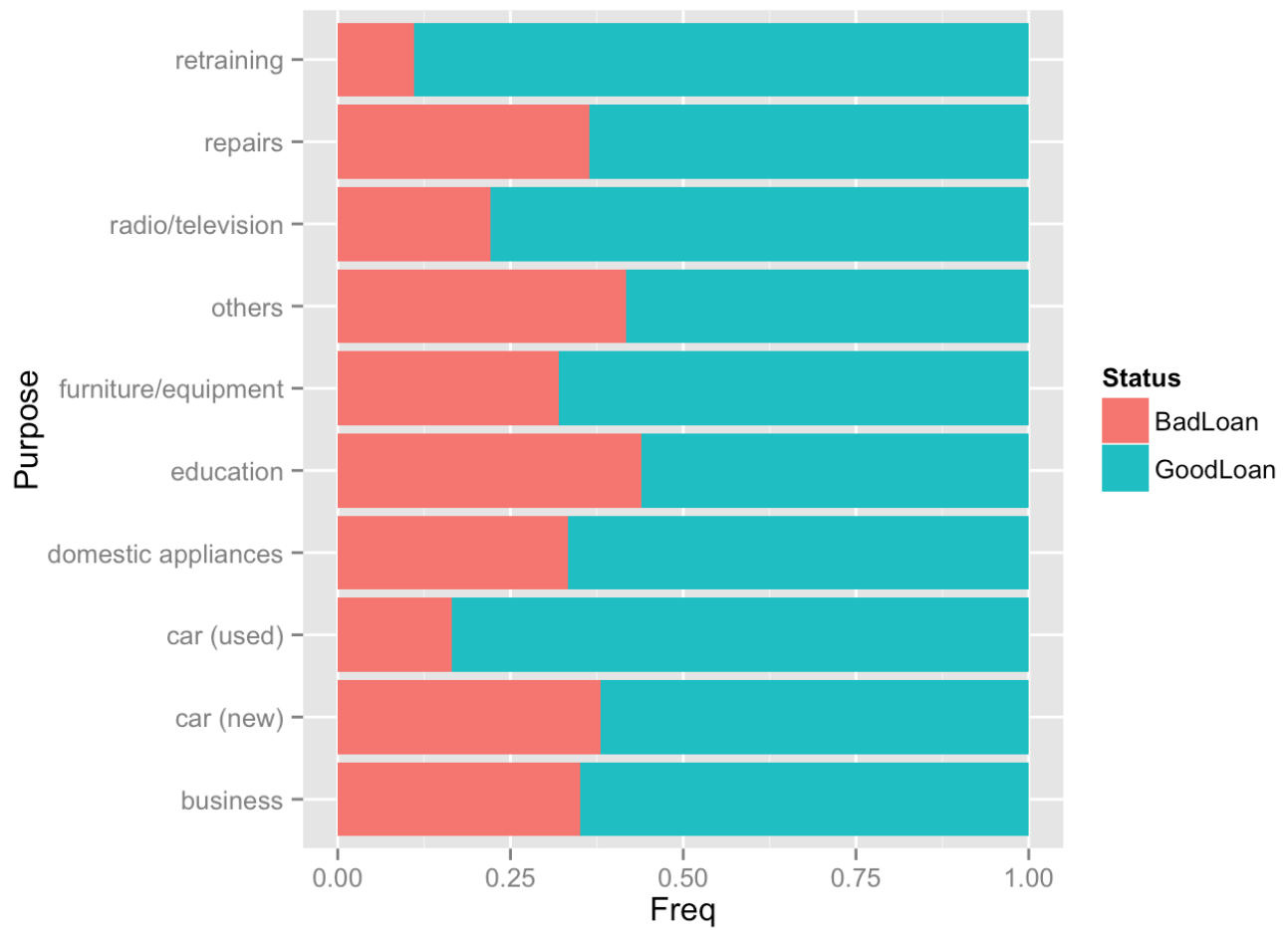


```
ggplot(tbl.df, aes(x=Purpose, y=Freq, fill=Status)) + geom_bar(stat="identity",  
position="fill")
```

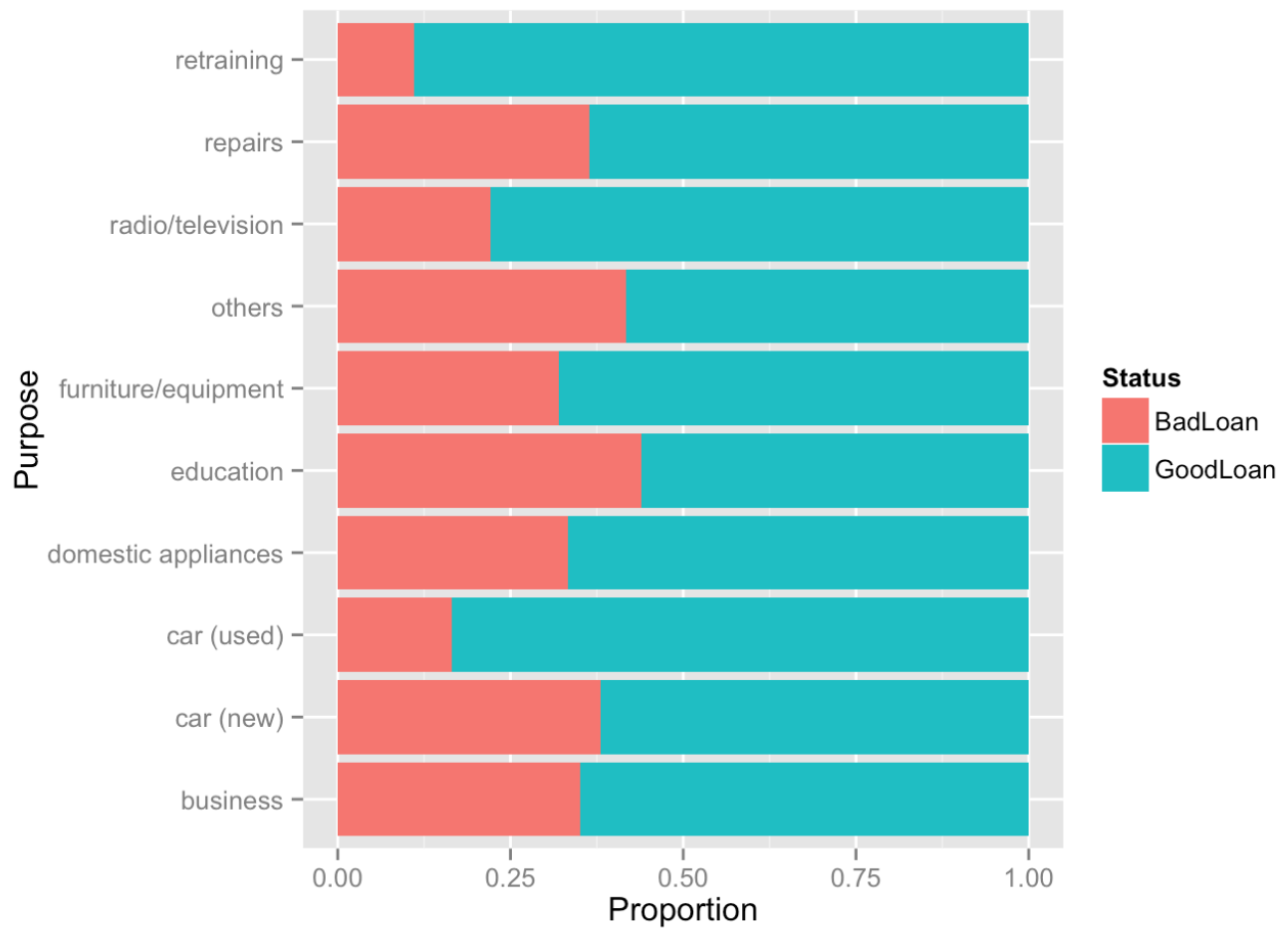




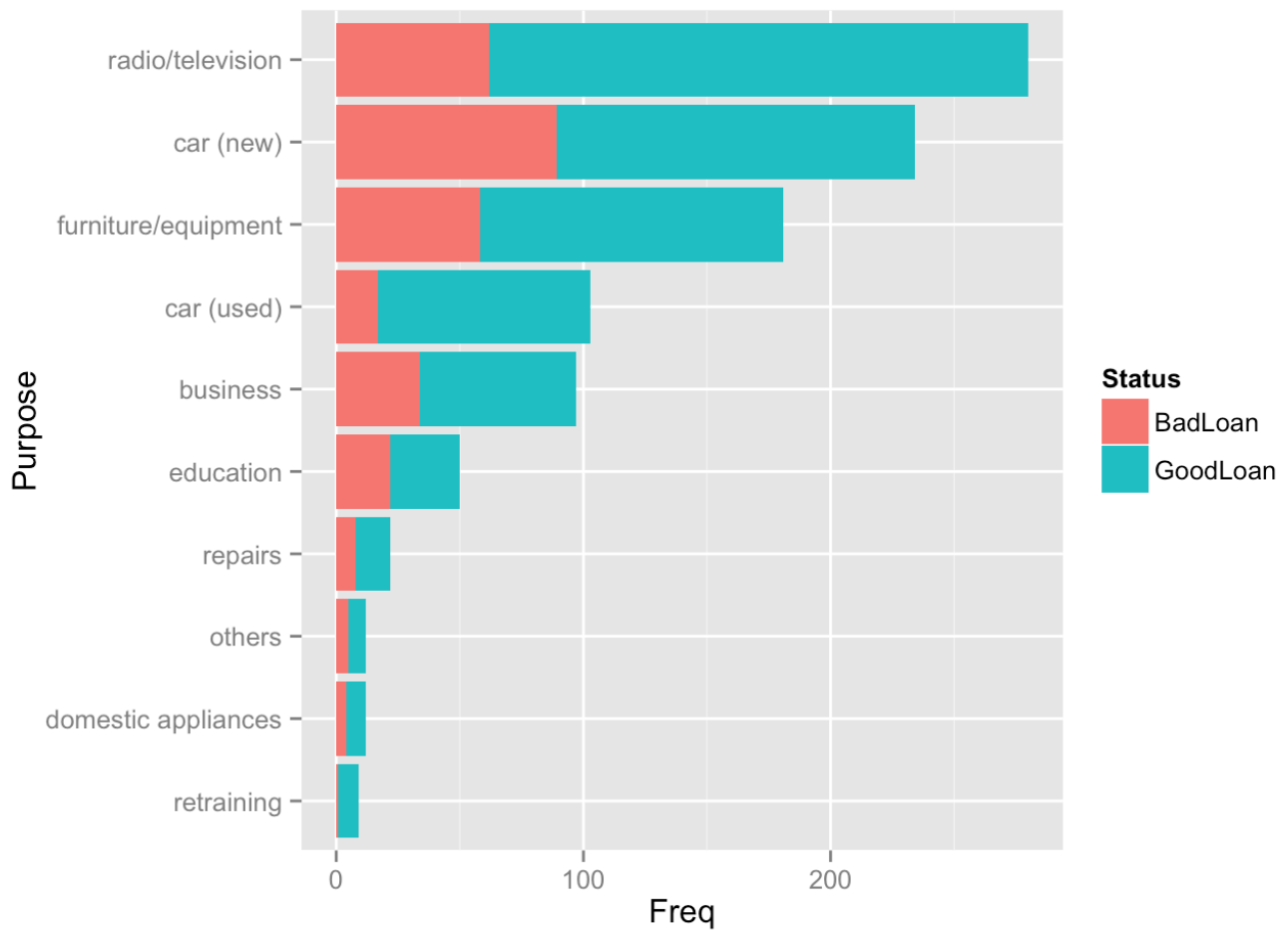
```
ggplot(tbl.df, aes(x=Purpose, y=Freq, fill=Status)) + geom_bar(stat="identity",
position="fill") +
  coord_flip()
```



```
ggplot(tbl.df, aes(x=Purpose, y=Freq, fill=Status)) + geom_bar(stat="identity",  
position="fill") +  
  coord_flip() +  
  xlab("Purpose") + ylab("Proportion")
```

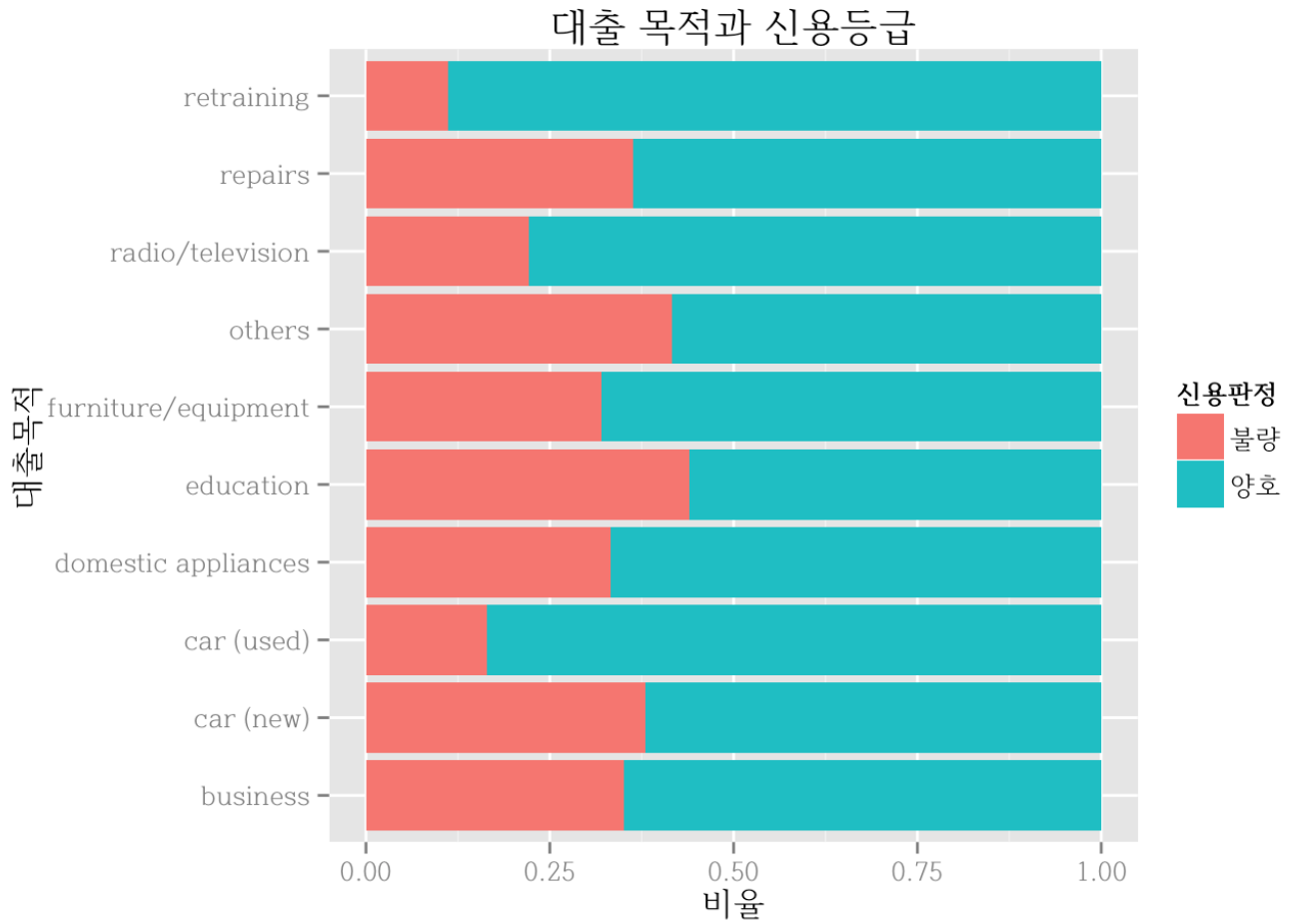


```
ggplot(tbl.df, aes(x=reorder(Purpose, Freq), y=Freq, fill=Status)) + geom_bar(s  
tat="identity") +  
  coord_flip() +  
  xlab("Purpose")
```



- 한글화

```
theme.kr <- theme(axis.title.x = element_text(family="HCR Batang LVT"),
axis.title.y = element_text(family="HCR Batang LVT"),
axis.text.x = element_text(family="HCR Batang LVT"),
axis.text.y = element_text(family="HCR Batang LVT"),
plot.title = element_text(family="HCR Batang LVT"),
legend.title = element_text(family="HCR Batang LVT"),
legend.text = element_text(family="HCR Batang LVT"))
ggplot(d) + geom_bar(aes(x=Purpose, fill=Good.Loan), position="fill") + coord_f
lip() +
  theme.kr +
  ylab("비율") + xlab("대출목적") +
  labs(fill="신용판정") +
  ggtitle("대출 목적과 신용등급") +
  scale_fill_discrete(labels=c("불량", "양호"))
```



```
ggplot(tbl.df, aes(x=reorder(Purpose, Freq), y=Freq, fill=Status)) + geom_bar(s
tat="identity") + coord_flip() +
  theme.kr +
  ylab("인원") + xlab("대출목적") +
  labs(fill="신용판정") +
  ggtitle("대출 목적과 신용등급") +
  scale_fill_discrete(labels=c("불량", "양호"))
```

## 대출 목적과 신용등급

대출 목적

